

Н.Д. Голев<sup>1</sup>

*Кемеровский государственный университет*

М.Е. Горюнова<sup>2</sup>

*Кемеровский государственный университет*

**ТЕКСТ КАК ОБЪЕКТ КВАНТИТАТИВНО-  
МОРФОЛОГИЧЕСКОГО ИССЛЕДОВАНИЯ**

В статье представлены результаты исследования текста, который рассматривался его авторами как предмет статистического расчета падежных форм в рамках проблемы противопоставления константных и неконстантных проявлений морфологических категорий в речевом материале. Выдвигается гипотеза о наличии русском речевом материале (корпусе текстов) тенденции к устойчивости квантитативного коэффициента каждой падежной формы. Авторы полагают, что данный расчет может послужить аналогом для других исследований такого типа и выступить материалом для идентификации и характеристики идиостиля автора. Работа проводится на материале трех художественных и одного научного текста. Инструментом расчета послужил Национальный корпус русского языка.

Ключевые слова: художественный текст, количественные методы, квантитативная морфология, национальный корпус русского языка, идиостиль, научный стиль, А.С.Пушкин, М.А.Булгаков, М.Ю.Лермонтов

N.D. Golev

*Kemerovo State University*

M.E. Goryunova

*Kemerovo State University*


**TEXT AS OBJECT OF QUANTITATIVE-MORPHOLOGICAL  
RESEARCH**

The article presents the results of a study of the text, which was considered by its authors as a subject of statistical calculation of case forms within the framework of the problem of contrasting constant and non-constant

---

<sup>1</sup> Николай Данилович Голев, доктор филологических наук, профессор кафедры русского языка Кемеровского государственного университета

<sup>2</sup> Мария Евгеньевна Горюнова, студент института филологии, иностранных языков и медиакоммуникаций Кемеровского государственного университета



manifestations of morphological categories in speech material. A hypothesis is advanced about the presence of Russian speech material (the body of texts) of the tendency to the stability of the quantitative coefficient of each case form. The authors believe that this calculation can serve as analogues for other studies of this type and serve as a material for identification and characterization of the author's idiostyle. The work is carried out on the basis of three artistic and one scientific text. The instrument of calculation was the National Corpus of the Russian language.

Key words: fiction, quantitative methods, quantitative morphology, national corpus of the Russian language, idiostyle, academic style, A.S. Pushkin, M.A. Bulgakov, M.Yu. Lermontov

Любой текст, став объектом научного исследования, предстает перед исследователем как сложная система, которая раскрывается и становится понятной при разных методах и подходах, отражающих не только поверхностные, но и глубинные характеристики текста. Одним из таких подходов является количественный метод лингвистического исследования. Возникший научный интерес к количественным характеристикам языка объяснил В.Г. Адмони: «Каждое языковое явление, помимо качественной определенности, говоря условно и образно, обладает и некоей «массой», т. е. занимает в кругу других языковых явлений некое пространство и соизмеримо с ними в неких измерениях с разной степенью интенсивности» [Адмони, 1963]. Количественные данные текста позволяют говорить о его содержательных специфических особенностях текста, различных особенностях его автора, стиля, что подтверждается рядом количественных исследований на разных уровнях языка: фонетическом [Булгакова, Шолохов, Томашенко, 2015], графическом [Петрова, Соловьева, 2016], лексическом [Ляшевская, Шаров, 2009], грамматическом [Герд, Капорулина, Колесов, 1974] и др.

Начало использования количественных приемов, связанное, как отмечает С.С. Верхозин, с появлением структурной лингвистики, в настоящее время переросло в востребованные методики исследования современной научной парадигмы. К примеру, без статистических данных не обходятся многие работы по социолингвистике, психолингвистике, стилеметрии и др.

Актуальность данных приемов можно объяснить характеристикой языка, которую обозначил Б.Н. Головин: 1) языку присущи количественные признаки, количественные характеристики, 2) существует связь между качественными и количественными характеристиками языковой системы, 3) частоты различных элементов языка в речевом потоке подчиняются статистическим законам [Головин, 1971]. Один из соавторов настоящей статьи также неоднократно задавался подобными вопросами, ср.: «Каким же образом может быть выявлены и зафиксированы количественные характеристики слова, которые далее могут быть достаточно надежно интерпретированы в качественно-семантическом аспекте?» [Голев, 2006, с. 10].

Однозначно можно говорить о количественно-качественной интерпретации текста в двух плоскостях: 1) идентификации, как конечной задаче квантитативного расчета каких-либо единиц текста. При этом может производиться идентификация как автора, так стиля и эпохи; 2) характеристика идиостиля автора, где статистические данные иллюстрируют, как текст данного автора проявляет свою сущность в каком-либо (фонетическом, лексическом, грамматическом) уровне языка и как та или иная языковая система проявляет себя, функционируя в тексте.

Идентификация автора с помощью статистических данных текста может иметь любую задачу направленность (культурную, правовую, лингвистическую и др.). Существующие исследования показывают, что языковые единицы любого уровня, проанализированные в тексте со стороны их статистической представленности, могут служить данными для атрибуции. Например, Борисов Л.А., Орлов Ю.Н. Осминин К.П. в работе «Идентификация автора текста по распределению частот буквосочетаний» предлагают исследовать текст и его принадлежность к какому-либо автору с помощью статистики распределения буквосочетаний [Борисов, Орлов, Осминин, 2013]. На лексическом уровне Г.В. Напреенко предложила лексико-квантитативные методы идентификации языковой личности [Напреенко, 2016], которые позволяют моделировать языковую

---

личность по индивидуальному набору лексико-квантитативных закономерностей.

Устойчивые характеристики употребления каких-либо единиц в текстах одного автора (а также стиля и эпохи) позволили говорить о них, как значимых проявлениях идиостиля автора. Стилеметрия, как дисциплина, изучающая стилевые характеристики с точки зрения их упорядоченности и систематизации, использует в качестве инструмента исследования различные вычислительные методы, в том числе и квантитативные. Рассматривается не только авторское употребление фонем, лексем, граммем, но и как уровни языковой системы проявляют себя в тексте.

Направления изучения текста со стороны статистических его данных подтверждают свою способность (как к идентификации, так и к характеристике идиостиля), как говорилось ранее, на разных уровнях языковой системы. Однако именно морфологическая система является наиболее трудным объектом для исследования квантитативных закономерностей. Трудность обуславливается сложным процессом предварительной обработки языковых данных, отсутствием однозначных методик для проведения статистического расчета. Несмотря на это, использование данных морфологического анализа текста для его толкования не является новым. Так, Л.В. Щерба в своем труде «Опыты лингвистического толкования стихотворений: «Воспоминание» Пушкина» обратил внимание на значение грамматических форм некоторых слов и их влияние на содержание произведения.

Например, «*умолкнет* и *наляжет* поставлены перед своими именительными падежами для достижения большей компактности соответственных фраз и для уменьшения их глагольности»; «...приименный родительный может быть поставлен перед своим определяемым. Дело в том, что приименный родительный, будучи поставлен после своего имени, склонен составлять особую группу, особенно если он имеет еще и свое определение»; «Слово шумный, хотя и является прилагательным от шум, однако в нашем сознании представляется производным от глагола шуметь, откуда и заимствует свою выразительность, являясь своего рода причастием в зародыше» и др. [Щерба, 1957]. Такие наблюдения иллюстрируют большие

возможности «квантитативной интуиции». Однако на современном этапе морфологический анализ текста предполагает не интуитивные, а точные методы.

Несколько предварительных замечаний о состоянии данного направления квантитативной морфологии. Механизация вычислительных процессов значительно упростила статистический расчет. На современном этапе, квантитативная морфология активно изучается на базе языковых корпусов. Корпусная лингвистика позволяет рассчитать ту или иную грамматическую форму в нужных текстах, нужного стиля, у нужного автора и т.д.. М.В. Копотев, исследуя статистику падежей на основе Национального корпуса русского языка (НКРЯ), отмечает, что «подкорпус со снятой омонимией в НКРЯ заслуживает доверия. Представляется, что при всех возможных неточностях составленная на его основе частотная грамматика может служить опорой в исследованиях» [Копотев, 2008] В своем исследовании мы опираемся на данные именно НКРЯ.

Национальный корпус русского языка – «это информационно-справочная система, основанная на собрании русских текстов в электронной форме. Корпус предназначен для всех, кто интересуется самыми разными вопросами, связанными с русским языком». Общий объем корпуса составляет 600 млн. слов, что позволяет выбрать определенный круг текстов для исследования. НКРЯ представляет возможность рассчитать все грамматические категории какого-либо текста. Все грамматические данные «в основном опираются на морфологическую модель, представленную в «Грамматическом словаре русского языка» А. А. Зализняка». Именно НКРЯ послужил инструментом для проведения настоящего квантитативного исследования.

Для точных результатов при работе в корпусе с морфологической системой мы избираем тексты со снятой грамматической омонимией. Материалом данного исследования послужили прозаические художественные тексты М.А. Булгакова («Мастер и Маргарита»), М.Ю. Лермонтова («Герой нашего времени»), А.С. Пушкина «Арап Петра Великого») и научный текст А. А. Антонова «Минералогия родингитов Баженовского гипербазитового массива». Сопоставление представленности

падежных форм заявленных текстов проиллюстрировано общим соотношением представленности категории падежа в тексте. Научный текст используется в исследовании как контрастивный материал.

Общая частотность употребления падежных форм в художественных произведениях выглядит следующим образом:

М.А. Булгаков «Мастер и Маргарита»

Именительный	35%
Винительный	20,8%
Родительный	16,3%
Творительный	9,8%
Предложный	9,2%
Дательный	5,9%

М.Ю. Лермонтов «Герой нашего времени»

Именительный	37,7%
Винительный	20,6%
Родительный	16,1%
Творительный	7,6%
Дательный	7,1%
Предложный	6,2%

А.С. Пушкин «Арап Петра Великого»:

Именительный	32,11%
Винительный	21,8%
Родительный	15,8%
Творительный	10,6%
Дательный	7,5%
Предложный	7%

Данные статистического расчета показывают, что в произведениях М.Ю. Лермонтова и А.С. Пушкина соотношение

падежей одинаково, однако количественные показатели не равны: Им.п. «Герой нашего времени» 37,7% / Им.п. «Арап Петра Великого» 32,11%; В.п. «Герой нашего времени» 20,6% / «Арап Петра Великого» 21,6%; Р.п. «Герой нашего времени» 16,1% / «Арап Петра Великого» 15, 8% и т.д. В романе М.А. Булгакова соотношение падежей отлично от произведений М.Ю.Лермонтова и А.С. Пушкина: наименее употребительным в «Мастере и Маргарите» является дательный падеж, предложный пятый по употребляемости, а в «Герое нашего времени» и «Арапе Петра Великого» наименее употребительным является предложный падеж, дательный пятый по частотности.

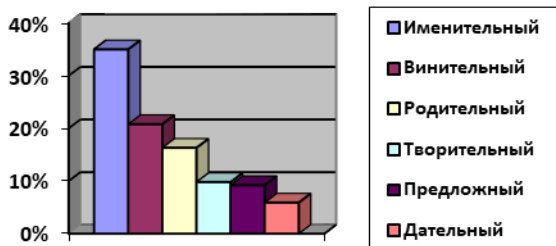
Одинаковое соотношение падежей, но различная (хоть и минимальная) количественная представленность, позволяют сделать некатегорические выводы, носящие вероятностный характер: 1) соотношение падежей говорит о такой закономерности распределения падежных форм, которую мы квалифицируем как проявление морфологической константы; 2) количественное расхождение подтверждает возможность идентификации и характеристики идиостиля по частотности падежных форм.

Частотность падежных форм в научном тексте выглядит так:

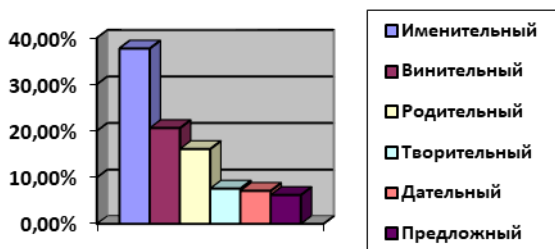
Соотношение падежных форм научного текста разительно отличается от данных художественного текста. Сходство минимальное, ограничивающееся творительным падежом, который и в художественных, и в научных текстах является четвертым по употребляемости. Для удобства сопоставления данные можно представить в следующих диаграммах:

Родительный	38,4%
Именительный	27,4%
Предложный	12,3%
Творительный	10,7%
Винительный	7,7%
Дательный	3,5%

### "Мастер и Маргарита"

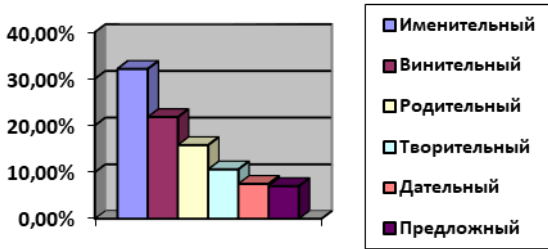


### "Герой нашего времени"

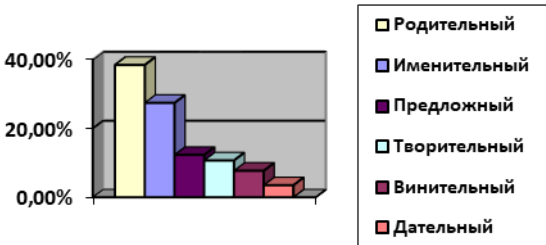




### "Арап Петра Великого"



### "Минералогия родингитов Баженовского гипербазитового массива"



Сопоставление данных количественного анализа падежных форм в трех текстах художественного стиля и одном научном показывает, что морфологическая система, в частности рассмотренная падежная система текста функционирует закономерно в текстах одного стиля (здесь – художественного), что подтверждает одинаковое (максимально приближенное) в количественном отношении распределение в произведениях частей речи, изменяющихся по падежам. Также закономерны «лидирующие» позиции некоторых падежей (именительный, винительный, родительный). Различие (т.е. отход от закономерности) позволяет анализировать морфологическую систему текста как составляющую

характеристику идиостиля автора. Текст другого стиля показывает и подтверждает гипотезу о том, что морфологическая система по-разному проявляет себя в функциональных стилях.

Таким образом, статистика падежных форм в художественных текстах, контрастирующая с представленностью падежа в научном тексте позволяет утверждать о наличии в русском языковом материале как тенденции к константности, так и тенденции к вариативности. Иными словами, тенденция к константности не является абсолютной. Что касается объема и спектра вариативности, то намеченная в статье статистика дает лишь начальный материал для продолжения исследований в данном направлении. Одно из них – направление обоснования – опровержения гипотезы о наличии в русском языковом материале морфологических универсалии для текстов одного стиля. Неконстантность, проявившаяся в некотором количественном расхождении частотности падежных форм в художественных текстах, дает основания для осторожного предположения о возможности использования данных количественного морфологического анализа как объекта изучения идиостиля того или иного автора. Другое направление прикладных количественных исследований – идентификационная лингвистика и лингво-экспертная (автороведческая) практика. Исследование в области морфологии может быть продолжено поиском константных-неконстантных грамматических признаков в текстах разных авторов, стилей, эпох и служить инструментом и материалом описания содержательных особенностей того или иного текста или группировок текстов, в том числе по признаку принадлежности одному автору. Глобальной перспективой количественной морфологии является расширение спектра статистических исследований – не только падежа, но и других грамматических категорий (вида, залога времени), различных частей речи, грамматических форм (причастия, деепричастий инфинитива). Таким путем количественная морфология имеет шанс выравняться с другими уровнями количественными лингвистическими дисциплинами по теоретическому моделированию русского текста и выхода в прикладную плоскость.

В этой связи предполагаем, что статистический расчет позволяет определить разность представления константности и неконстантности на разных языковых уровнях: фонетическом [Булгакова, Шолохов, Томашенко, 2015], графическом [Петрова, Соловьева, 2016], лексическом [Ляшевская, Шаров, 2009; Тамбовцев, Тамбовцева, Тамбовцева, 2008; Напреенко, 2016], грамматическом [Герд, Капорулина, Колесов, 1974] и др. Предварительное рассмотрение результатов данных исследований позволяет сформулировать гипотезу, согласно которой - чем «выше» языковой уровень, тем слабее тенденция к константности и сильнее тенденция к вариативности. В другом аспекте - чем больше возможности выбора языковой единицы у автора, тем сильнее проявления неконстантности и – соответственно - возможности функциональной интерпретации квантитативных различий разных текстов.

#### СПИСОК ЛИТЕРАТУРЫ

**Адмони, В.Г.** Качественный и количественный анализ грамматических явлений / В.Г. Адмони // Вопросы языкознания. – 1963. – № 4. – С. 57–63.

**Борисов Л.А., Орлов Ю.Н., Осминин К.П.** Идентификация автора текста по распределению частот буквосочетаний // Препринты ИПМ им. М.В.Келдыша. 2013. № 27. 26 с. URL: <http://library.keldysh.ru/preprint.asp?id=2013-27>

**Булгакова Е.В., Шолохов А.В., Томашенко Н.А.** Метод идентификации дикторов на основе сравнения статистик длительностей фонем // Научно-технический вестник информационных технологий, механики и оптики. № 1 (95). Санкт-Петербург, 2015. С. 70-77.

**Верхозин С.С.** О статусе количественных методов в лингвистике // Вестник иркутского государственного лингвистического университета. № 3 (24). Иркутск, 2013. С. 145-150.

**Герд А.С., Капорулина Л.В., Колесов Е.В., Черепанова О.А., Рускова М.П.** Именное склонение в славянских языках XI–XIV вв.: Лингвостатистический анализ. Санкт-Петербург, 1974

**Голев Н.Д.** О новых источниках квантитативного речевого материала и их семасиологических возможностях (постановка

---

проблем // Изменяющаяся Россия: новые парадигмы и новые решения в лингвистике: Материалы I Международной конференции (Кемерово, 29-31 августа 2006 г.): В 4-х частях. Часть 4. - Кемерово: Юнити, 2006. – С. 9-15.

**Голев Н.Д., Напреенко Г.В.** Формально-количественное моделирование языковой личности: идентификационный аспект // Языковая личность: моделирование, типология, портретирование // Сибирская лингвоперсонология Москва, 2016. С. 76-93.

**Головин Б.Н.** Язык и статистика. / Б.Н. Головин // М., 1971.

**Ляшевская О.Н., Шаров С.А.** Частотный словарь современного русского языка (на материалах Национального корпуса русского языка). М.: Азбуковник, 2009.

**Петрова А.Л., Соловьева А.А.** Исследование распределений букв, передающих гласные, в итальянских, русских, французских и немецких песенных текстах методами математической статистики // «Математика и междисциплинарные исследования – 2016» Сборник докладов всероссийской научно-практической конференции молодых ученых с международным участием. гл. ред. Ю.А. Шарапов; Пермский государственный национальный исследовательский университет. Пермь, 2016. С. 261-265.

**Тамбовцев Ю.А., Тамбовцева А.Ю., Тамбовцева Л.А.** Типология распределения некоторых лингвистических единиц в тексте как показатель авторства текста // Вестник Омского государственного университета № 2. Омск, 2008. С. 88-96.

**Щерба Л.В.** Опыты лингвистического толкования стихотворений: «Воспоминание» Пушкина / Л.В. Щерба // Избранные работы по русскому языку. Москва.: Учпедгиз, 1957. С. 26–44.

**Коптев М.В.** К построению частотной грамматики русского языка: падежная система по корпусным данным // Slavica Helsingiensia (Инструментарий русистики: корпусные подходы). Вып. 34. Helsinki, 2008. S. 136–151.

**References:**

**Admoni, V.G.** Kachestvennyj i kolichestvennyj analiz grammaticheskikh yavlenij / V.G. Admoni // Voprosy yazykoznanija. – 1963. – № 4. – S. 57–63. (In Russian)

**Borisov L.A., Orlov YU.N., Osminin K.P.** Identifikaciya avtora teksta po raspredeleniyu chastot bukvosochetaniy // Preprinty IPM im. M.V.Keldysha. 2013. № 27. 26 s. URL: <http://library.keldysh.ru/preprint.asp?id=2013-27> (In Russian)

**Bulgakova E.V., Sholohov A.V., Tomashenko N.A.** Metod identifikacii diktorov na osnove sravneniya statistik dlitel'nostej fonem // Nauchno-tehnicheskij vestnik informacionnyh tekhnologij, mekhaniki i optiki. № 1 (95). Sankt-Peterburg, 2015. S. 70-77. (In Russian)

**Verhozin S.S.** O statuse kolichestvennyh metodov v lingvistike // Vestnik irkutskogo gosudarstvennogo lingvisticheskogo universiteta. № 3 (24). Irkutsk, 2013. S. 145-150. (In Russian)

**Gerd A.S., Kaporulina L.V., Kolesov E.V., Cherepanova O.A., Ruskova M.P.** Imennoe sklonenie v slavyanskikh yazykah XI–XIV vv.: Lingvostaticeskij analiz. Sankt-Peterburg, 1974 (In Russian)

**Golev N.D.** O novyh istochnikah kvantitativnogo rechevogo materiala i ih semasiologicheskikh vozmozhnostyakh (postanovka problem // Izmenyayushchayasya Rossiya: novye paradigmy i novye resheniya v lingvistike: Materialy I Mezhdunarodnoj konferencii (Kemerovo, 29-31 avgusta 2006 g.): V 4-h chastyah. CHast' 4. - Kemerovo: YUniti, 2006. – S. 9-15. (In Russian)

**Golev N.D., Napreenko G.V.** Formal'no-kolichestvennoe modelirovanie yazykovoj lichnosti: identifikacionnyj aspekt // YAzykovaya lichnost': modelirovanie, tipologiya, portretirovanie // Sibirskaya lingvopersonologiya Moskva, 2016. S. 76-93. (In Russian)

**Golovin B.N.** YAzyk i statistika. / B.N. Golovin // M., 1971. (In Russian)

**Lyashevskaya O.N., Sharov S.A.** CHastotnyj slovar' sovremennogo russkogo yazyka (na materialah Nacional'nogo korpusa russkogo yazyka). M.: Azbukovnik, 2009. (In Russian)

**Petrova A.L., Solov'eva A.A.** Issledovanie raspredelenij bukvy, peredayushchih glasnye, v ital'yanskikh, russkikh, francuzskikh i nemeckikh pesennyh tekstah metodami matematicheskoy statistiki // «Matematika i

---

mezhdisciplinarnye issledovaniya – 2016» Sbornik dokladov vsrossijskoj nauchno-prakticheskoj konferencii molodyh uchenyh s mezhdunarodnym uchastiem. gl. red. YU.A. SHarapov; Permskij gosudarstvennyj nacional'nyj issledovatel'skij universitet. Perm', 2016. S. 261-265. (In Russian)

**Tambovcev YU.A., Tambovceva A.YU., Tambovceva L.A.** Tipologiya raspredeleniya nekotoryh lingvisticheskikh edinic v tekste kak pokazatel' avtorstva teksta // Vestnik Omskogo gosudarstvennogo universiteta № 2. Omsk, 2008. S. 88-96. (In Russian)

**Shcherba L.V.** Opyty lingvisticheskogo tolkovaniya stihotvorenij: «Vospominanie» Pushkina / L.V. SHCHerba // Izbrannye raboty po russkomu yazyku. Moskva.: Uchpedgiz, 1957. S. 26–44. (In Russian)

**Kopotev M.V.** K postroeniyu chastotnoj grammatiki russkogo yazyka: padezhnaya sistema po korpusnym dannym // Slavica Helsingiensia (Instrumentarij rusistiki: korpusnye podhody). Vyp. 34. Helsinki, 2008. S. 136–151. (In Russian)